BIOMARKERS AND ASSAYS FOR CARCINOGENESIS

This application is a continuation-in-part of US provisional application Serial No. 60/118,078, filed on January 29, 1999, the contents of which are hereby incorporated herein.

Field of the Invention

The present invention relates to genes differentially regulated by phenobarbital, nucleic acid molecules or fragments thereof that act as biomarkers for carcinogenesis, and nucleic acid molecules that are useful as probes or primers for detecting or inducing carcinogenesis, respectively. The invention also relates to applications such as forming antibodies capable of binding carcinogenesis biomarkers or fragments thereof.

Background

In the field of toxicology, high resolution assays now make it possible to discover differences in gene expression brought on by exposure to a particular xenobiotic. Such high-throughput, high-resolution molecular biology methods can be used to determine virtually all toxicant-induced changes in gene expression. A catalog of toxicant-induced gene expression changes would be useful to better predict animal toxicity in order to reduce costs, timelines, and animal use by enhancing the probability that product candidates chosen for further development will pass regulatory testing requirements. Such a catalog would also enable scientists to better predict human toxicity, resulting in fewer compounds failing in clinical trials while better safeguarding human health.

The basis for these types of investigations is the expectation that toxicological endpoints (e.g. tumor formation) are the result of earlier molecular events. For example, by creating a catalog of changes in rat liver gene expression following treatment with phenobarbital, one can test whether early gene expression

is as predictive as later readouts in assessing the nongenotoxic carcinogenicity of this compound in rats.

The power of transcriptional genomic analyses is that they can measure changes in the expression of thousands of genes, and a comprehensive catalog of expression changes can be envisioned. Using the same catalog of changes, other known nongenotoxic carcinogens (NGCs) could be assessed, as well as compounds known not to be NGCs in rats. Analysis of correlations between the changes and carcinogenesis, as well as analysis of the biological significance of the genes, should indicate whether there are specific genes or gene-expression patterns that predict carcinogenesis. Thus, there is a need in the art for catalogs or panels of predictive markers. Such panels of expressed genes would allow one to examine a greater number of candidate compounds in a shorter period of time prior to selecting a lead compound for traditional testing. As a result of this screening approach, the success rate of compounds in pre-clinical trials should improve dramatically.

These panels of predictive markers could also be used to assess the use of primary rat hepatocytes in high-throughput cell-based assays of toxicity and carcinogenicity. This would further increase the number of compounds that could be assessed, perhaps to the point where entire compound libraries could be assayed, and scores for potential toxicities could be created for each compound. Further, parallel analyses using both animal and human genes could be used to correlate the results from pre-clinical in vivo and in vitro data (using both cultured animal and cultured human cells) with human clinical data to create assays that better predict human toxicity.

Summary Of The Invention

It is an object of the present invention to provide a catalog or panel of changes in gene expression that are predictive of carcinogenicity. The catalog

includes substantially-purified nucleic acid sequences that have been discovered. In one embodiment, the present invention relates to a substantially-purified nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ NO: 1 through SEQ NO: 580 or fragments, substantial homologues, and substantial complements thereof.

In another embodiment, the present invention relates to a substantially-purified carcinogenesis biomarker or fragment thereof encoded by a first nucleic acid molecule which substantially hybridizes to a second nucleic acid molecule, the second nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ NO:1 through SEQ NO:580 and complements thereof.

It is another object of the present invention to provide an assay for toxicity to predict the carcinogenicity of a composition. In a further embodiment, the present invention relates to a method for measuring the carcinogenicity of a composition comprising exposing a mammal to the composition; and determining the presence or absence of mRNA which substantially hybridizes to a nucleic acid sequence selected from the group consisting of SEQ NO:1 through SEQ NO:580 and complements thereof.

It is a further object of the present invention to provide a quantitative and qualitative method of detection of carcinogenesis-related proteins or peptides of the present invention. In one embodiment, antibodies, proteins, peptides, or fusion proteins that specifically bind to one or more of the proteins encoded by the nucleic acid molecules of the present invention can be used to measure the cacinogenesis-related proteins.

Various other objects and advantages of the present invention will become apparent from the following figures and description of the invention.

Brief Description of the Drawings

Figure 1 shows a comparison of mRNA levels of differentially expressed transcripts.

Detailed Description Of The Invention

A. General Concepts and Definitions

These detailed descriptions are presented for illustrative purposes only and are not intended as a restriction on the scope of the invention. Rather, they are merely some of the embodiments that one skilled in the art would understand from the entire contents of this disclosure. All parts are by weight and temperatures are in Degrees centigrade unless otherwise indicated.

Abbreviations and Definitions

The following is a list of abbreviations and the corresponding meanings as used interchangeably herein:

IMDM = Iscove's modified Dulbecco's media

mg = milligram

ml or mL = milliliter

μg or ug= microgram

μl or ul = microliter

ODNs= oligonucleotides

PCR= polymerase chain reaction

RP-HPLC = reverse phase high performance liquid chromatography

The following is a list definitions of various terms used herein:

The term "**altered**" means that expression differs from the expression response of cells or tissues not exhibiting the phenotype.

The term "**amino acid(s)**" means all naturally occurring L-amino acids.

The term "**biologically active**" means activity with respect to either a structural or a catalytic attribute, which includes the capacity of a nucleic acid to hybridize to another nucleic acid molecule, or the ability of a protein to be bound by an antibody (or to compete with another molecule for such binding), among others. Catalytic attributes involve the capacity of the agent to mediate a chemical reaction or response.

The term "**cluster**" means that BLAST scores from pairwise sequence comparisons of the member clones are similar enough to be considered identical with experimental error.

The term "**complement**" means that one nucleic acid exhibits complete complementarity with another nucleic acid.

The term "**complementarity**" means that two molecules can hybridize to one another with sufficient stability to permit them to remain annealed to one another under conventional high stingency conditions.

The term "**complete complementarity**" means that every nucleotide of one molecule is complementary to a nucleotide of another molecule.

The term "**degenerate**" means that two nucleic acid molecules encode for the same amino acid sequences but comprise different nucleotide sequences (see US Patent 4,757,006).

The term "**exogenous genetic material**" means any genetic material, whether naturally occurring or otherwise, from any source that is capable of being inserted into any organism.

The term "**expression response**" means the mutation affecting the level or pattern of the expression encoded in part or whole by one or more nucleic acid molecules.

The term "**fragment**" means a nucleic acid molecule whose sequence is shorter than the target or identified nucleic acid molecule and having the identical, the

substantial complement, or the substantial homologue of at least 7 contiguous nucleotides of the target or identified nucleic acid molecule.

The term "**fusion protein**" means a protein or fragment thereof that comprises one or more additional peptide regions not derived from that protein. Such molecules may be derivatized to contain carbohydrate or other moieties (such as keyhole limpet hemocyanin, etc.).

The term "**hybridization probe**" means any nucleic acid capable of being labeled and forming a double-stranded structure with another nucleic acid over a region large enough for the double stranded structure to be detected.

The term "**isolated**" means an agent is separated from another specific component with which it occurred. For example, the isolate material may be purified to essential homogeneity, as determined by PAGE or column chromatography, such as HPLC. An isolated nucleic acid can comprise at least about 50, 80, or 90% (on a molar basis) of all macromolecular species present. Some of these methods described later lead to degrees of purification appropriate to identify single bands in electrophoresis gels. However, this degree of purification is not required.

The term "**marker nucleic acid**" means a nucleic acid molecule that is utilized to determine an attribute or feature (*e.g.*, presence or absence, location, correlation, etc.) of a molecule, cell, or tissue.

The term "**mimetic**" refers to a compound having similar functional and/or structural properties to another known compound or a particular fragment of that known compound.

The term "**minimum complementarity**" means that two molecules can hybridize to one another with sufficient stability to permit them to remain annealed to one another under at least conventional low stringency conditions.

The term "**PCR probe**" means a nucleic acid capable of initiating a polymerase activity while in a double-stranded structure with another nucleic acid. For

example, Krzesicki, *et al.*, *Am. J. Respir. Cell Mol. Biol.* 16:693-701 (1997), incorporated by reference in its entirety, discusses the preparation of PCR probes for use in identifying nucleic acids of osteoarthrits tissue. Other methods for determining the structure of PCR probes and PCR techniques have been described.

The term "**phenotype**" means any of one or more characteristics of an organism, tissue, or cell.

The term "**polymorphism**" means a variation or difference in the sequence of the gene or its flanking regions that arises in some of the members of a species.

The term "**primer**" means a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (e.g., in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but must be sufficiently complementary to hybridize with a template.

The term "**probe**" means an agent that is utilized to determine an attribute or feature (e.g. presence or absence, location, correlation, etc.) of a molecule, cell, tissue, or organism.

The term "**product score**" refers to a formula which indicates the strength of a BLAST match using the fraction of overlap of two sequences and the percent identitiy. The formula is as follows:

$$\text{Product Score} = \frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq1)}, \text{length(Seq2)}\}}$$

The term "**promoter region**" means a region of a nucleic acid that is capable, when located in *cis* to a nucleic acid sequence that encodes for a protein or peptide, of functioning in a way that directs expression of one or more mRNA molecules.

The term "**protein fragment**" means a peptide or polypeptide molecule whose amino acid sequence comprises a subset of the amino acid sequence of that protein.

The term "**protein molecule/peptide molecule**" means any molecule that comprises five or more amino acids.

The term "**recombinant**" means any agent (e.g., DNA, peptide, etc.), that is, or results from, however indirectly, human manipulation of a nucleic acid molecule. The recombination may occur inside a cell or in a tube.

The term "**selectable marker**" means a gene who's expression can be detected by a probe as a means of identifying or selecting for transformed cells.

The term "**specifically bind**" means that the binding of an antibody or peptide is not competitively inhibited by the presence of non-related molecules.

The term "**specifically hybridizing**" means that two nucleic acid molecules are capable of forming an anti-parallel, double-stranded nucleic acid structure.

The term "**substantial complement**" means that a nucleic acid sequence shares at least 80% sequence identity with the complement.

The term "**substantial fragment**" means a fragment which comprises at least 100 nucleotides.

The term "**substantial homologue**" means that a nucleic acid molecule shares at least 80% sequence identity with another.

The term "**substantial identity**" means that 70% to about 99% of a region or fragment in a molecule is identical to a region of a different molecule. When the individual units (e.g., nucleotides or amino acids) of the two molecules are schematically positioned to exhibit the highest number of units in the same position over a specific region, a percentage identity of the units identical over the total

number of units in the region is determined. Numerous algorithmic and computerized means for determining a percentage identity are known in the art. These means may allow for gaps in the region being considered in order to produce the highest percentage identity.

The term "**substantially hybridizes**" means that two nucleic acid molecules can form an anti-parallel, double-stranded nucleic acid structure under conditions (e.g. salt and temperature) that permit hybridization of sequences that exhibit 90% sequence identity or greater with each other and exhibit this identity for at least a contiguous 50 nucleotides of the nucleic acid molecules.

The term "**substantially purified**" means that one or more molecules that are or may be present in a naturally occurring preparation containing the target molecule will have been removed or reduced in concentration.

Agents of the Invention

A. Nucleic Acid Molecules

The present invention relates to nucleic acid sequences selected from the group consisting of SEQ NO:1 through SEQ NO: 580, substantial fragments thereof, substantial homologues thereof, and substantial complements thereof. By creating a catalog of changes in rat liver gene expression following treatment with phenobarbital, substantially-purified nucleic acid sequences selected from the group consisting of SEQ NO: 1 through SEQ NO: 580 have been discovered. These sequences are useful as biomarkers of carcinogenesis.

The present invention also relates to nucleic acid sequences derived from the one or more sequences identified in SEQ NOS:1-580. Fragment nucleic acids may encompass significant portion(s) of, or indeed most of, these sequences. For example, a fragment nucleic acid can encompass an carcinogenesis biomarker gene homolog or fragment thereof. Alternatively, the fragments may comprise smaller

oligonucleotides, for example an oligonucleotide having from about 10 to about 250 nucleotides or from about 15 to about 30 nucleotide.

A variety of computerized means for identifying sequences derived from the SEQ NO.: 1-580 exists. These include the five implementations of BLAST, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN), as well as FASTA and others (Coulson, *Trends in Biotechnology* 12:76-80 (1994); Birren *et al.*, *Genome Analysis* 1:543-559 (1997)). Other programs which use either individual sequences or make models from related sequences to further identify sequences derived from SEQ NO 1- SEQ NO 580 exist. Model building and searching programs includes HMMer (Eddy), MEME (Bailey and Elkan, *Ismb* 3: 21-29 (1995)) and PSI-BLAST (Altschul et al., *Nucleic Acids Res* 25: 3389-3402 (1997)). Another set of programs which use predicted, related, or known protein structures to further identify sequences derived from SEQ NO 1- SEQ NO 580 exists. Structure-based searching programs includes ORF and PROSITE. Other programs which use individual sequences or related groups of sequences relying on pattern discovery to further identify sequences derived from SEQ NO:1-580 exist. Pattern recognition programs include Teiresias (Rigoutsos, I. and A. Floratos, *Bioinformatics* 1: (1998)). These programs can search any appropriate database, such as GenBank, dbEST, EMBL, SwissProt, PIR, and GENES. Furthermore, computerized means for designing modifications in protein structure are also known in the art (Dahiyat and Mayo, *Science* 278:82-87 (1997)).

Nucleic acids or fragments thereof of the present invention are capable of specifically hybridizing to other nucleic acids under certain circumstances. The present invention further relates to nucleic acid sequences that will specifically hybridize to one or more of the nucleic acids set forth in SEQ NO: 1 through SEQ NO: 580, or complements thereof, under moderately stringent conditions, for

example at about 2.0 X SSC and about 65°C. Alternatively, the nucleic acid sequences of the present invention may specifically hybridize to one or more of the nucleic acids set forth in SEQ NO:1 through SEQ NO: 580, or complements thereof, under high stringency conditions.

The present invention also relates to nucleic acid sequences that share between 100% and 90% sequence identity with one or more of the nucleic acid sequences set forth in SEQ NO: 1 through to SEQ NO: 580 or complements thereof. In a further aspect of the invention, nucleic acid sequences of the invention share between 100% and 95% sequence identity with one or more of the nucleic acid sequences set forth in SEQ NO: 1 through SEQ NO: 580, or complements thereof. Alternatively, nucleic acid sequences of the present invention may share between 100% and 98% or between 100% and 99% sequence identity with one or more of the nucleic acid sequences set forth in SEQ NO: 1 through SEQ NO: 580, or complements thereof.

A region or fragment in a molecule with "substantial identity" to a region of a different molecule can be represented by a ratio. In a preferred embodiment, a 10 nucleotide in length nucleic acid region or fragment of the invention has a percentage identity of about 70% to about 99% with a nucleic acid sequence existing within one of SEQ NO.: 1-580 or a complement of SEQ NO.: 1-580.

The invention also provides a computer-readable medium having recorded thereon the sequence information of one or more of SEQ NO:1 through SEQ NO:580, or complements thereof. In addition, the invention provides a method of identifying a nucleic acid comprising providing a computer-readable medium of the invention and comparing nucleotide sequence information using computerized means.

i. Nucleic Acid Primers and Probes

The present invention also relates to nucleic acid primers and probes derived from the nucleic acid sequences set forth in SEQ NO: 1 through SEQ NO: 580. The nucleic acid primers and probes of the invention may be derived from the disclosed sequences, such as a fragment of 10 nucleotides or more or a sequence with 70% to 99% identity to a fragment of at least 10 nucleotides. Numerous methods for defining or identifying primers and probes for nucleic acid or sequence based analysis exist. Examples of suitable primers include, but are not limited to, the nucleic acid sequences set forth in SEQ NO: 519 through SEQ NO: 580. Examples of 5' primers (from the 5' to 3' direction) include, but are not limited to, SEQ NO: 550-580. Examples of 3' primers (from the 5' to 3' direction) include, but are not limited to, SEQ NO: 519-549. Examples of suitable probes include, but are not limited to, the nucleic acid sequences set forth in SEQ NO: 490 through SEQ NO: 518. The genes that corresponds to the primer and probe sequences (SEQ NO: 490-580) are described in Table 7.

Conventional stringency conditions are described by Sambrook, *et al.*, *Molecular Cloning, A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York (1989), and by Haymes, *et al. Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985), the entirety of both is herein incorporated by reference. Departures from complete complementarity are therefore permissible, as long as such departures do not completely preclude the capacity of the molecules to form a double-stranded structure. Thus, in order for a nucleic acid molecule to serve as a primer or probe it need only be sufficiently complementary in sequence to be able to form a stable double-stranded structure under the particular solvent and salt concentrations employed.

Appropriate stringency conditions that promote DNA hybridization, for example, 6.0 X sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 X SSC at 50°C, are known to those skilled in the art or can be found in Ausubel, *et al.*, *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989) (see especially sections 6.3.1-6.3.6). [This reference and the supplements through January 2000 are specifically incorporated herein by reference and can be relied to make or use any embodiment of the invention.] For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 X SSC at 50°C to a high stringency of about 0.2 X SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C. Temperature and salt conditions may be varied independently.

Primers and probes of the present invention can be used in hybridization assays or techniques, in a variety of PCR-type methods, or in computer-based searches of databases containing biological information. Exemplary methods include a method of identifying a nucleic acid which comprises the hybridization of a probe of the invention with a sample containing nucleic acid and the detection of stable hybrid nucleic acid molecules. Also included are methods of identifying a nucleic acid comprising contacting a PCR probe of the invention with a sample containing nucleic acid and producing multiple copies of a nucleic acid that hybridizes, or is at least minimally complementary, to the PCR probe.

The primers and probes of the invention may be labeled with reagents that facilitate detection (e.g., fluorescent labels, Prober et al., Science 238: 336-340 (1987), Albarella et al., EP 144914;, chemical labels, Sheldon et al., U.S. Patent 4,582,789, Albarella et al., U.S. Patent 4,563,417; and modified bases, Miyoshi et al., EP 119448) all of which are incorporated by reference in their entirety)).

ii. Nucleic Acids Comprising Genes, Fragments, or Homologs Thereof

This invention also provides genes corresponding to the cDNA sequences disclosed herein, also called carcinogenesis biomarkers. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. The methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials.

In another preferred embodiment, nucleic acid molecules having SEQ NO: 1 through SEQ NO: 580, or complements and fragments of either, can be utilized to obtain homologues equivalent to the naturally existing homologues.

In a further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding a homologue or fragment thereof in SEQ NO: 1 through SEQ NO: 580, or complements thereof, due to the degeneracy in the genetic code in that they encode the same protein but differ in nucleic acid sequence. In another further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding an homologue of fragment thereof in SEQ NO: 1 through SEQ NO:580, or complements thereof, due to fact that the different nucleic acid sequence encodes a protein having one or more conservative amino acid residue. Examples of conservative substitutions are set forth below. Codons capable of coding for such conservative substitutions are well known in the art.

Original Residue	Conservative Substitutions
Ala	ser
Arg	lys
Asn	gln; his
Asp	glu
Cys	ser; ala
Gln	asn
Glu	asp
Gly	pro
His	asn; gln
Ile	leu; val
Leu	ile; val
Lys	arg; gln; glu
Met	leu; ile
Phe	met; leu; tyr
Ser	thr
Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

Genomic sequences can be screened for the presence of protein homologues utilizing one or a number of different search algorithms have that been developed, such as the suite of BLAST programs. The BLASTX program allows the comparison of nucleic acid sequences in this invention to protein databases.

In a preferred embodiment of the present invention, the homologue protein or fragment thereof exhibits a BLASTX probability score of less than 1E-30,

alternatively a BLASTX probability score of between about 1E-30 and about 1E-12 or a BLASTX probability score of greater than 1E-12 with a nucleic acid or gene of this invention. In another preferred embodiment of the present invention, the nucleic acid molecule encoding the gene homologue or fragment thereof exhibits a % identity with its homologue of between about 25% and about 40%, or alternatively between about 40% and about 70%, or from 70% and about 90%, or from about 90% and 99%. In another embodiment, the gene homologue or fragment has a single nucleotide difference from its homologue.

The resulting product score of a BLAST program ranges from 0 to 100, with 100 indicating 100% identity over the entire length of the shorter of the two sequences, and 0 representing no shared identity between the sequences. The homologue protein or fragment thereof may also exhibit a product score of 100. Alternatively, the product score is between about 49 and about 99. The protein or fragment may also exhibit a product score of 0. Alternatively, the homolog or fragment exhibits a product score between about 1 and about 49.

The sequences of the present invention were searched for sequence similarity and given biological annotations based on that similarity.

Table 1: Sequences down-regulated at least 1.7-fold by 13 weeks of treatment with phenobarbitol are shown with their corresponding annotation.

Table 2: Sequences up-regulated at least 1.7-fold by 13 weeks of treatment with phenobarbitol are shown with their corresponding annotation.

Table 3: Sequences down-regulated at least 1.7-fold by 5 weeks of treatment with phenobarbitol are shown with their corresponding annotation.

Table 4: Sequences upregulated at least 1.7-fold by 5 weeks of treatment with phenobarbitol are shown with their corresponding annotation.

iv. Vectors and Host Cells Containing Nucleic Acid Molecules

The present invention also relates to recombinant DNA molecules comprising a nucleic acid sequence of the invention and a vector. The invention further relates to host cells (mammalian and insect) that containing the recombinant DNA molecules. Methods for obtaining such recombinant mammalian host cell, comprising introducing exogenous genetic material into a mammalian host cell are also provided by the invention. The present invention also relates to an insect cell comprising a mammalian cell containing a mammalian recombinant vector. The present invention also relates to methods for obtaining a recombinant mammalian host cell, comprising introducing into a mammalian cell exogenous genetic material.

A recombinant protein may be produced by opererably linking a regulatory control sequence to a nucleic acid of the present invention and putting it into an expression vector. Regulatory sequences include promoters, enhancers, and other expression control elements which are described in Goeddel (*Hene Expression Technology: Methods in Enzymology* 185. Academic Press, San Diego, CA (1990)). For example, the native regulatory sequences or regulatory sequences native to the transformed host cell can be used. One of skill in the art is familiar with numerous examples of these additional functional sequences, as well as other functional sequences, that may optionally be included in an expression vector. The design of the expression vector may depend on such factors as the choice of the host cell to be transformed, and/or the type of protein desired. Many such vectors are commercially available, including linear or enclosed elements (see for example, Broach, et al., Experimental Manipulation of Gene Expression, ed. M. Inouye, Academic Press, (1983); Sambrook, *et al., Molecular Cloning, A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York (1989)). Typically, expression constructs will contain one or more selectable

markers, including the gene that encodes dihydrofolate reductase and the genes that confer resistance to neomycin, tetracycline, ampicillin, chloramphenicol, kanamycin and streptomycin resistance.

Prokaryotic and eukaryotic host cells transfected by the described vectors are also provided by this invention. For instance, cells which can be transfected with the vectors of the present invention include, but are not limited to, bacterial cells such as *E. coli* (e.g., *E. coli* K 12 strains), *Streptomyces*, *Pseudomonas*, *Serratia marcescens* and *Salmonella typhimurium*, insect cells (baculovirus), including *Drosophila*, fungal cells, such as yeast cells, plant cells, and ovary cells (CHO), and COS cells.

One may use different promoter sequences, enhancer sequences, or other sequences which will allow for enhanced levels of expression in the expression host. Thus, one may combine an enhancer from one source, a promoter region from another source, a 5'- noncoding region upstream from the initiation methionine from the same or different source as the other sequences, and the like. One may provide for an intron in the non-coding region with appropriate splice sites or for an alternative 3'- untranslated sequence or polyadenylation site. Depending upon the particular purpose of the modification, any of these sequences may be introduced, as desired.

Where selection is intended, the sequence to be integrated will have an associated marker gene, which allows for selection. The marker gene may conveniently be downstream from the target gene and may include resistance to a cytotoxic agent, e.g. antibiotics, heavy metals, resistance or susceptibility to HAT, gancyclovir, etc., complementation to an auxotrophic host, particularly by using an auxotrophic yeast as the host for the subject manipulations, or the like. The marker gene may also be on a separate DNA molecule, particularly with primary mammalian cells. Alternatively, one may screen the various transformants, due to

the high efficiency of recombination in yeast, by using hybridization analysis, PCR, sequencing, or the like.

For homologous recombination, constructs can be prepared where the amplifiable gene will be flanked, normally on both sides, with DNA homologous with the DNA of the target region. Depending upon the nature of the integrating DNA and the purpose of the integration, the homologous DNA will generally be within 100 kb, usually 50 kb, preferably about 25 kb, of the transcribed region of the target gene, more preferably within 2 kb of the target gene. Where modeling of the gene is intended, homology will usually be present proximal to the site of the mutation. The term gene is intended to encompass the coding region and those sequences required for transcription of a mature mRNA. The homologous DNA may include the 5'-upstream region outside of the transcriptional regulatory region, or comprise any enhancer sequences, transcriptional initiation sequences, adjacent sequences, or the like. The homologous region may include a portion of the coding region, where the coding region may be comprised only of an open reading frame or combination of exons and introns. The homologous region may comprise all or a portion of an intron, where all or a portion of one or more exons may also be present. Alternatively, the homologous region may comprise the 3'-region, so as to comprise all or a portion of the transcriptional termination region, or the region 3' of this position. The homologous regions may extend over all or a portion of the target gene or be outside the target gene comprising all or a portion of the transcriptional regulatory regions and/or the structural gene.

Thus, the nucleic acid molecules described can be used to produce a recombinant form of the protein via microbial or eukaryotic cellular processes. Ligating the polynucleic acid molecule into a gene construct, such as an expression vector, and transforming or transfecting into hosts, either eukaryotic (yeast, avian, insect, plant, or mammalian) or prokaryotic (bacterial cells), are standard

procedures used in producing other well known proteins. Similar procedures, or modifications thereof, can be employed to prepare recombinant proteins according to the present invention by microbial means or tissue-culture technology. Accordingly, the invention pertains to the production of encoded proteins or polypeptides by recombinant technologies.

B. Proteins and Polypeptides

The present invention also relates to proteins, peptides and polypeptides encoded by the nucleic acid sequences of the invention. Protein and peptide molecules can be identified using known protein or peptide molecules as a target sequence or target motif in the BLAST programs of the present invention. These proteins, peptides and poylpeptides of the invention can be made using the nucleic acids or derived from the sequence information of the nucleic acids are also disclosed in the present invention. This invention also provides a compound or composition comprising one or more polypeptides, which comprise: 1) at least one fragment, segment, or domain of at least 15-1,000 contiguous amino acids, with at least one portion encoded by one or more of SEQ NOS: 1-580; 2) at least one amino acid sequence selected from those encoding at least one of SEQ NOS: 1-580; or 3) at least one modification corresponding to fragments, segments, or domains within one of SEQ NOS:1- 580. The proteins, peptides and polypeptides of the invention can be made recombinantly as described above. Alternatively, the proteins, peptides and polypeptides of the invention can be produced synthetically.

Protein fragments or fusion proteins may be derivatized to contain carbohydrate or other moieties (such as keyhold limpet hemocyanin, etc.). A fusion protein or peptide molecule of the present invention is preferably produced via recombinant means.

Modifications can be naturally provided or deliberately engineered into the nucleic acids, proteins, and polypeptides of the invention to generate variants. For example, modifications in the peptide or DNA sequences can be made by those skilled in the art using known techniques, such as site-directed mutagenesis. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of one or more selected amino acid residues. For example, one or more cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Additional cysteine residues can also be added as a substitute at sites to promote disulfide bonding and increase stability. Techniques for identifying the sites for alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art. Techniques for making alterations, substitutions, replacements, insertions or deletions (see, e.g., U.S. Pat. No. 4,518,584) are also well known in the art. Preferably, any modification of a protein, polypeptide, or nucleic acid of the invention will retain at least one of the structural or functional attributes of the molecule.

The polypeptide or protein can also be tagged to facilitate purification, such as with histidine- or methionine-rich regions [His-Tag; available from LifeTechnologies Inc, Gaithersvurg, MD] that bind to metal ion affinity chromatography columns, or with an epitope that binds to a specific antibody [Flag, available from Kodak, New Haven, CT].

A number of purification methods or means are also known and can be used. For example, reverse-phase high performance liquid chromatography (RP-HPLC).

C. Antibodies

This invention also provides an antibody, polyclonal or monoclonal, that specifically binds at least one epitope found in or specific to a carcinogenesis

biomarker protein or polypeptide or a protein or polypeptide, of fragment or variant thereof, of this invention. Antibodies can be generated by recombinant, synthetic, or hybridoma technologies. One aspect of the present invention concerns antibodies, single-chain antigen binding molecules, or other proteins that specifically bind to one or more of the protein or peptide molecules of the present invention and their homologues, fusions or fragments. Such antibodies may be used to quantitatively or qualitatively detect the protein or peptide molecules of the present invention.

Nucleic acid molecules that encode all or part of the protein of the present invention can be expressed, by recombinant means, to yield protein or peptides that can in turn be used to elicit antibodies that are capable of binding the expressed protein or peptide. Such antibodies may be used in immunoassays for that protein or peptide. Such protein-encoding molecules or their fragments may be a "fusion" molecule (*i.e.*, a part of a larger nucleic acid molecule) such that, upon expression, a fusion protein is produced. It is understood that any of the nucleic acid molecules of the present invention may be expressed, by recombinant means, to yield proteins or peptides encoded by these nucleic acid molecules.

The antibodies that specifically bind proteins and protein fragments of the present invention may be polyclonal or monoclonal, and may comprise intact immunoglobulins, or antigen binding portions of immunoglobulins (such as (F(ab'), $F(ab')_2$ fragments), or single-chain immunoglobulins producible, for example, via recombinant means. Conditions and procedures for the construction, manipulation and isolation of antibodies (see, for example, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, New York (1988), the entirety of which is herein incorporated by reference) are well known in the art.

As discussed below, such antibody molecules or their fragments may be used for diagnostic purposes. Where the antibodies are intended for diagnostic purposes, it may be desirable to derivatize them, for example with a ligand group (such as biotin) or a detectable marker group (such as a fluorescent group, a radioisotope or an enzyme).

The ability to produce antibodies that bind the protein or peptide molecules of the present invention permits the identification of mimetic compounds of those molecules. Combinatorial chemistry techniques, for example, can be used to produce libraries of peptides (see WO 9700267), polyketides (see WO 960968), peptide analogues (see WO 9635781, WO 9635122, and WO 9640732), oligonucleotides for use as mimetic compounds derived from this invention. Mimetic compounds and libraries can also be generated through recombinant DNA-derived techniques. For example, phage display libraries (see WO 9709436), DNA shuffling (see US Patent 5,811,238) other directed or random mutagenesis techniques can produce libraries of expressed mimetic compounds. It is understood that any of the agents of the present invention can be substantially purified and/or be biologically active and/or recombinant.

Uses of the Invention

The present invention also provides methods for identifying carcinogen compounds. The nucleic acids, peptides and proteins of the invention can be useful in predicting the toxicity of test compounds. Nucleic acids represent biomarkers which are correlated to an altered cellular state. These markers, individually or in combination, can be measured in response to compounds to screen for those compounds that suppress or activate the genes and thus alter the state of the cell in an undesired manner. Specifically, the nucleic acids, peptides and proteins can be used directly in numerous methods well known in the art to identify or detect the presence of specific nucleic acid or amino acid sequences.

Carcinogens can be identified by contacting an animal, tissue from a mammal, or a mammalian cell, such as a rat hepatocyte, with a compound, under conditions allowing production of mRNA by the cell. The resulting mRNA is then separated and its presence or absence detected. Differential expression of these biomarkers can be monitored in tissues and fluids at the mRNA level using methods well known in the art such as Northern hybridizations, RNAase protection, NMR, rt-PCR, and *in situ* hybridizations. *In vitro* techniques can also be used to detect differential expression of genomic DNA such as, for example, Southern hybridizations.

Similarly, differential expression of these biomarkers can be monitored at the protein level using, for example, enzyme linked immunosorbent assays (ELISAs), Western blots, HPLC-liquid chromotography, NMR, immunoprecipitations and immunofluorescence. Protein identification can also be performed using new techniques including biomolecular interaction analysis (BIA) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF). (Nelson *et al.*, Interfacing biomolecular interaction analysis with mass spectrometry and the use of bioreactive mass spectrometer probe tips in protein characterization, in Techniques in Protein Chemistry VIII, p. 493-504, 1997; Kalrsson *et al.*, Experimental design for kinetic analysis of protein-protein interactions with surface plasmon resonance biosensors, J. Immun. Meth, 220, 121-133, 1997; Krone *et al.*, BIA/MS: Interacting biomolecular interaction analysis with mass spectrometry, Anal. Chem. 244, 124-132, 1997; and Wong *et al.*, Validation parameters for a novel biosensor assay which simultaneously measures serum concentrations of a humanized monoclonal antibody and detects induced antibodies, J. Immun. Meth, 209, 1-15, 1997.)

Using the catalog of the present invention, one skilled in the art can predict with the tested compound is a carcinogen. Compounds that results in the

production of nucleic acids, peptides or protein from the catalog, or a subset of catalog, are carcinogenic. To be able to predict carcinogenic, one need not use all of the nucleic acids or peptides of the present invention. For example, if one tested for all of the disclosed biomarkers and found 20% or more to be differentially expressed this would predict that the test compound is a carcinogen. Alternatively, one could use a sub-set of the biomarkers, such as, for example, 20-30 of the nucleic acids. With such a sub-set one would expect 70-80% to be differentially expressed when the test compound is a carcinogen. In addition, one could select only a few of the biomarkers, for example, 10, and look for 100% of them to be differentially expressed as an indication of a carcinogen.

mRNA, protein, or genomic DNA of the invention can be detected in biological samples including, for example, tissues, cells, or biological fluids from a subject such as blood, urine, or liver and thyroid tissue.

Various microarrays, beads, glass or nylon slides, membranes or other repeatable assay apparati can be constructed using the nucleic acids, peptides, and proteins of the present invention. These apparati can then be used to detect differential expression of these biomarkers. A non-limiting description of selected methods follows.

A. Microarrays

In one embodiment, the nucleic acids of the invention can be used to monitor expression. A microarray-based method for high-throughput monitoring of gene expression may be utilized to measure carcinogenesis biomarker hybridization targets. This 'chip'-based approach involves using microarrays of nucleic acids as specific hybridization targets to quantitatively measure expression of the corresponding genes (Schena *et al.*, *Science 270*:467-470 (1995), the entirety of which is herein incorporated by reference; Shalon, Ph.D. Thesis, Stanford University (1996), the entirety of which is herein incorporated by reference). Every

nucleotide in a large sequence can be queried at the same time. Hybridization can also be used to efficiently analyze nucleotide sequences.

Several microarray methods have been described. One method compares the sequences to be analyzed by hybridization to a set of oligonucleotides or cDNA molecules representing all possible subsequences (Bains and Smith, *J. Theor. Biol. 135*:303 (1989), the entirety of which is herein incorporated by reference). A second method hybridizes the sample to an array of oligonucleotide or cDNA probes. An array consisting of oligonucleotides or cDNA molecules complementary to subsequences of a target sequence can be used to determine the identity of a target sequence, measure its amount, and detect differences between the target and a reference sequence. Nucleic acid microarrays may also be screened with protein molecules or fragments thereof to determine nucleic acids that specifically bind protein molecules or fragments thereof.

The microarray approach may also be used with polypeptide targets (*see*, U.S. Patent Nos. 5,800,992, 5,445,934; 5,143,854, 5,079,600, 4,923,901, all of which are herein incorporated by reference in their entirety). Essentially, polypeptides are synthesized on a substrate (microarray) and these polypeptides can be screened with either protein molecules or fragments thereof or nucleic acid molecules in order to screen for either protein molecules or fragments thereof or nucleic acid molecules that specifically bind the target polypeptides (Fodor *et al.*, *Science 251*:767-773 (1991), the entirety of which is herein incorporated by reference).

B. Hybridization Assays

Oligonucleotide probes, whose sequences are complementary to that of a portion of the nucleic acids of the invention, such as SEQ NO.:1-580, can be constructed. These probes are then incubated with cell extracts of a patient under conditions sufficient to permit nucleic acid hybridization. The detection of double-

stranded probe-mRNA hybrid molecules is indicative of biomarkers of carcinogenesis or sequences derived from rat liver hepatocytes treated with a nongenotoxic carcinogen. Thus, such probes may be used to ascertain the level and extent of carcinogenesis or the production of certain proteins. The nucleic acid hybridization may be conducted under quantitative conditions or as a qualitative assay.

C. PCR Assays

A nucleic acid of the invention, such as one of SEQ NO.:1-580 or complements thereof, can be analyzed for use as a PCR probe. A search of databases indicates the presence of regions within that nucleic acid that have high and low regions of identity to other sequences in the database. Ideally, a PCR probe will have high identity with only the sequence from which it is derived. In that way, only the desired sequence is amplified. Computer generated searches using programs such as MIT Primer3 (Rozen and Skaletsky (1996, 1997, 1998)) , or GeneUp (Pesole, *et al.*, *BioTechniques 25*:112-123 (1998)), for example, can be used to identify potential PCR primers.

The PCR probes or primers can be used in methods such as described in Krzesicki, *et al.*, *Am. J. Respir. Cell Mol. Biol. 16*:693-701 (1997) (incorporated by reference in its entirety) to identify or detect sequences expressed in carcinogenesis.

These detailed descriptions are presented for illustrative purposes only and are not intended as a restriction on the scope of the invention. Rather, they are merely some of the embodiments that one skilled in the art would understand from the entire contents of this disclosure. All parts are by weight and temperatures are in Degrees centigrade unless otherwise indicated.

EXAMPLES

The following examples will illustrate the invention in greater detail, although it will be understood that the invention is not limited to these specific examples. Various other examples will be apparent to the person skilled in the art after reading the present disclosure without departing from the spirit and scope of the invention. It is intended that all such other examples be included within the scope of the appended claims.

Example 1

Rats were treated with phenobarbital for thirteen weeks or in a separate experiment, for 5 days. Liver mRNAs were extracted and probed for those mRNAs specifically altered by phenobarbital treatment by comparing with mRNA expression in untreated rats. The relative abundance of cellular mRNAs in rat liver was determined using PE GenScope's AFLP (Amplified Fragment Length Polymorphism)-based Transcript Imaging technology. The mRNA is converted into double-stranded cDNA, which is then cut with restriction enzymes. The resulting restriction fragments are tagged with specific adapters of known sequences, which allows for subsequent amplification of the fragments under highly stringent conditions. Similar technology has been used in plants (Money, T. et al., Nucleic Acids Res. 24:2616-2617 (1996), incorporated by reference in its entirety).

Specifically, rats were treated by oral gavage for 88 days in the 13 week experiment, or for 5 days with 200 mg/kg phenobarbital or control vehicle. The average expression levels of mRNAs for three phenobarbital-induced genes (P450 2B1, P450 3A1, and UDP-glucuronosyl transferase) were measured using RT-PCR, and showed substantial induction of mRNA expression levels as compared to control rats.

In one study, ten differentially expressed transcript derived fragments (TDF's) were isolated and cloned. For each TDF, four or five colonies were picked and their sequences determined using standard sequencing techniques. In each case, all colonies sequenced contained the same sequences. This is a reflection of the ability to reduce the complexity of the AFLP gel profile by using primers with additional selective nucleotides. The ten TDF sequences were BLASTed against GenBank. The identities of the bands were consistent with what one might predict would be altered by treatment with phenobarbital. PCR analysis of the samples confirmed that these genes are differentially expressed following treatment.

Example 2

Validation of AFLP Biomarkers by rt-PCR (Taqman)

After AFLP experiments were conducted, and results analyzed, the effects of phenobarbital on the expression of several biomarkers were validated. RNA was extracted from the same liver samples used in the AFLP study, in addition to liver samples from rats treated with phenobarbital for 2-weeks, followed by reverse transcription reactions to generate cDNA, followed by PCR, using Taqman technology. The genes analyzed for phenobarbital-induced alterations, and the corresponding AFLP sequence numbers are listed in Table 5, and a graph and a chart of the actual results are in Table 6 and Figure 1.

The results indicate that AFLP technology can find biomarkers. Eleven of the 17 (65%) genes analyzed were also determined to be differentially expressed using rt-PCR. However, this is based on comparisons at the same timepoint (13 weeks). When the rt-PCR analyses performed on the 2 week samples are considered, another marker (S-033) is found to be differentially expressed. Theoretically, differences in sensitivity and/or specificity between the two techniques could be accounted for these minor discrepancies. However, S-033 is an

example of how AFLP has identified biomarkers which are optimal for carcinogen detection at timepoints other than 13 weeks.

As noted above, the specific examples should not be interpreted as a limitation to the scope of the invention. Instead, they are merely exemplary embodiments one skilled in the art would understand from the entire disclosure of this invention.

TABLE 1

SEQ NO	Annotation*
275	rat mRNA for (S)-2-hydroxy acid oxidase
276	human NADH-ubiquinone oxidoreductase
277	rat mRNA organic anion transporter 3
278	Ula-1 RNA from transformed mouse cell line
279	rat hemoglobin alpha chain gene
280	rat mRNA for calcium binding protein
281	rat heat shock protein 27
282	rat mRNA for 50-kDa bone sialic acid
283	rat mRNA for lactate dehydrogenase
284	rat ribonuclease 4 mRNA
285	mouse Src-associated adaptor protein
286	rat mRNA for plasminogen protein
287	rat gene 33 DNA
288	rat mRNA for 50-kDa bone sialic acid
289	mouse glycolate oxidase mRNA
290	rat mRNA for cytochrome b5
291	mouse mRNA for tripeptidyl peptidase II
292	human eukaryotic protein synthesis init.
293	rat fatty liver acid binding protein
294	rat mRNA for ATP-stimulated glucocorticoid receptor translocation promoter
295	mouse apolipoprotein A-I/CIII mRNA
296	rat fibronectin (cell-, heparin-, and fibrin-binding domains)
297	rat mRNA encoding liver fatty acid binding
298	rat RoBo-1 mRNA
299	rat mRNA for pre-alpha-inhibitor, heavy chain
300	rat pancreatic secretory trypsin inhibitor
301	rat apolipoprotein A-IV mRNA
302	rat apolipoprotein A-IV mRNA
303	rat lecithin: cholesterol acyltransferase
304	mouse mRNA for very-long-chain acyl-CoA
305	rat Cyp3a locus
306	rat gene for alpha-fibrinogen
307	mouse protein phosphatase-1 binding protein
308	novel human mRNA similar to rat 45 kDa secretory protein
309	
310	rat retinol dehydrogenase type III mRNA
311	rat mRNA for lecithin-cholesterol acyltransferase
312	rat oxidative 17 beta hydroxysteroid dehydrogenase
313	rat hydroxysteroid sulfotransferase mRNA
314	mouse major histocompatibility locus cla
315	mouse ubiquitinating enzyme E2-230 kDA mRNA
316	mouse fatty acid transport protein 5 mRNA

317 rat (TSC-22) mRNA
318 rat SMP30 mRNA for senescence marker protein

TABLE 2

SEQ NO	Annotation
319	rat cytochrome P450
320	rat cytochrome P450b
321	rat cytochrome P450
322	
323	rat cytochrome P450 mRNA, 3' end
324	rat mRNA for carboxylesterase precursor
325	rat cytochrome P450e
326	rat aldehyde dehydrogenase (ALDH) mRNA
327	rat mRNA for carboxylesterase precursor
328	rat aldehyde dehyrdogenase (ALDH) mRNA
329	rat lipoprotein lipase mRNA
330	rat cytochrome P450IIB3
331	rat mRNA for P450IIIA23 protein
332	rat aflatoxin B1 aldehyde reductase
333	rat ,RMA for cytochrome P450 3A
334	rat testosterone 6-beta-hydroxylase (CYP 3A1) mRNA
335	rat mRNA for amyloidogenic glycoprotein
336	rat cytochrome P50 PB1 (PB1 allele) mRNA
337	rat epoxide hydrolase mRNA
338	rat mRNA for P450IIIA23 protein
339	rat CYP 3A1 mRNA
340	rat mRNA for hydroxysteroid sulfotransferase
341	rat mRNA for cytochrome P450
342	rat NADPH-cytochrome P450 reductase mRNA
343	
344	rat liver glutathione-S-transferase Yb-1
345	rat cytochrome P450 processed pseudogene
346	rat mRNA for glutathione S-transferase
347	rat NADPH-cytochrome P450 reductase mRNA
348	rat mRNA for P450IIIA23 protein
349	rat delta-aminolevulinate synthase mRNA
350	rat mRNA for glutathione S-transferase
351	rat mRNA for amyloidogenic glycoprotein
352	human GSTT1 mRNA
353	rat cytochrome P450IIB3
354	rat mRNA for glutathione transferase subunit 8
355	rat cytochrome P450IIB3
356	rat NADPH-cytochrome P450 reductase mRNA
357	rat glutathione S-transferase mRNA
358	rat NADPH-cytochrome P450 oxidoreductase
359	mouse mRNA for glutathione S-transferase
360	glutathione S-transferase
361	rat mRNA for glutathione transferase subunit 8

362	rat NADPH-cytochrome P450 oxidoreductase
363	rat cytochrome P450 PB1 (PB1 allele) mRNA
364	rat cytochrome P450 PB1 (PB1 allele) mRNA
365	glutathione S-transferase Yc1 subunit
366	rat 5-aminolevulinate synthase mRNA
367	rat cytochrome P450f mRNA
368	rat mRNA for polyubiquitin, 5' end
369	M. aureus mRNA for cytochrome P450IIC
370	preprocathepsin B (mouse, B16a melanoma)
371	rat phosphoglucomutase mRNA
372	rat malic enzyme gene, exon 4
373	rat mRNA for gluthathione S-transferase
374	rat cytochrome P450 mRNA
375	rat cytochrome P450 mRNA
376	rat cytochrome P450 mRNA
377	
378	human mitochondrial prostatein C3 subunit homolog
379	rat cytochrome P450 3A9 mRNA
380	rat cytochrome P450-1/PB- (ps) gene, exon
381	rat Hsp70-1 gene
382	rat cytochrome P450 mRNA
383	
384	human mRNA for transcription factor BTF
385	mesocricetus auratus mRNA for carboxylesterase
386	rat aromatic L-amino acid decarboxylase
387	rat mRNA for putative progesterone binding protein
388	rat Y-b3 glutathione S-transferase mRNA
389	rat NADPH-cytochrome P450 reductase mRNA
390	rat cytochrome PB23 mRNA
391	UGT2B4, UDP-glucuronosyltransferase 2B4
392	rat glutathione S-transferase A3 subunit
393	rat mRNA for cytochrome b5
394	rat mRNA for glutathione S-transferase
395	rat cytochrome P450 3A9 mRNA
396	glutathione s-transferase Yc1 subunit
397	bilirubin-specific UDP-glucuronosyltransferase
398	rat cytochrome P450 mRNA
399	rat p450Md mRNA for cytochrome P450
400	mouse glutathione S-transferase class mu
401	
402	
403	rat mRNA for beta-tubulin T beta15
404	human micosomal glutathione s-transferase
405	rat transketolase mRNA
406	rat cytochrome P450 (female-specific and growth hormone-inducible) mRNA

407	rat cytochrome P450 (female-specific and growth hormone-inducible) mRNA
408	NPT4, sodium phosphate transporter
409	rah- ras-related homolog (mouse, HT4 neuro)
410	human mRNA for 16G2
411	rat mRNA for analicular multidrug resistance
412	rat UDP-glucuronosyltransferase UGT1A7 mRNA
413	human sodium phosphate transporter (NPT4)
414	rat liver apolipoprotein A-I mRNA
415	rat UDP-glucuronosyltransferase mRNA
416	rat apolipoprotein A-I gene
417	mouse gene encoding tetranectin
418	mouse COP9 complex subunit 7a (COPS7a) mRNA

TABLE 3

SEQ NO	Annotation
419	rat mRNA for hydroxysteroid sulfotransferase
420	Zfp-29 gene for zinc finger protein
421	human HFREP-1 mRNA
422	mouse ATP sulfurylase/APS kinase 2
423	
424	mouse secreted apoptosis-related protein
425	human zinc finger gene ZNF2
426	rat angiotensinogen (PAT) gene, exon 2
427	
428	mouse methyltransferase (Cyt19)
429	mouse activin beta-c precursor gene
430	
431	
432	
433	
434	rat mRNA for hepatic lipase
435	
436	human (H326) mRNA
437	human mRNA for KIAA00181 gene
438	
439	mouse mRNA for paladin gene
440	
441	mouse activin beta-c precursor gene
442	rat orphan receptor RLD-1 (rld-1) mRNA
443	mouse oncomodulin gene (exon 1)
444	rat kallistatin mRNA mRNA
445	
446	rat gonadotropin-releasing hormone
447	URP- nuclear calmodulin-binding protein gb113vrtp
448	mouse Jun co-activator Jab1 (Jab 1) mRNA
449	rat zinc finger binding protein mRNA
450	mouse inhibitor of apoptosis protein 2 mRNA
451	
452	rat mRNA for glutathione peroxidase I
453	mouse CRBPI mRNA for cellular retinol
454	mouse wagneri mRNA for heat shock
455	mouse NPC1 (Npc1) mRNA
456	
457	

TABLE 4

SEQ NO	Annotation
458	rat UDP-glucuronosyltransferase-2 (UDPGT)
459	rat ribosomal protein S12 mRNA
460	rat ornithine decarboxylase (ODC) mRNA
461	rat cytokeratin 8 polypeptide mRNA
462	rat mRNA for cathepsin L
463	human rho GDI mRNA
464	rat CLP36 (clp36) mRNA
465	annexin II, 36 kDa calcium-dependent phos.
466	
467	rat ribosomal protein S18 mRNA
468	rat ornithine decarboxylase (ODC) mRNA
469	mouse (C57BL/6) GB-like mRNA
470	cyclic protein-2, cathepsin L proenzyme
471	human p27 mRNA
472	rat c-myc oncogene and flanking regions
473	rat mRNA for canalicular multispecific
474	mouse ctla-2-beta mRNA homolog
475	rat 3-hydroxy-3-methlyglutaryl CoA reductase
476	rat stathmin mRNA
477	rat mRNA for Mx1 protein
478	
479	rat mRNA for protein phosphatase-2A catalytic subunit
480	rat mRNA for Mx2 protein
481	human mRNA for MUF1 protein
482	mouse MA-3 (apoptosis-related gene) mRNA
483	human BRCA2 region, mRNA sequence CG012
484	
485	pre-mtHSP70, 70 kDa heat shock protein
486	
487	house mouse mRNA for MAP kinase, kinase 3B
488	rat mRNA for 14-3-3 protein gamma-subtype, putative protein kinase C
489	human homolog of the Aspergillus nidulans sudD gene product

* ANNOTATIONS REPRESENT THE PREDICTION OF THE BIOLOGICAL FUNCTIONS OF THE SEQUENCES BASED ON SIMILARITY TO KNOWN SEQUENCES.

TABLE 5

SEQ. NO.	Gene
3	Rat P-450
4	Rat aldehyde dehydrogenase
6	Rat UDPGT1.1
10	Rat vitamin D-binding protein
179	Rat UDPGT
25	Rat cytochrome B
114	Rat delta-aminolevulinate synthase
129	Glutathione S-transferase
34	Rat liver catalase
38	Rat alpha-2u globulin
40	Rat NADP-dep.isocitrate dehydrogenase
42	Mouse JAK1 (protein tyrosine kinase)
230	Rat carboxylesterase
46	Rat cathepsin B
52	(s)-2-hydroxy acid oxidase
116	Estrogen sulfotransferase
92	Rat nicotinic receptor alpha 7 subunit

TABLE 6

SEQ	Fold Change		
NO.	2-week	13-week	AFLP
3	1.34	1.85	2.3
4	16.36	12.88	8.2
6	0.93	1.5	4.6
10	0.66	0.79	1.7
179	14.11	9.05	10.5
25	1.85	0.75	4.2
114	1.22	4.03	3.8
129	2.52	4.03	4
34	0.79	0.45	1.6
38	0.35	0.03	0.04
40	0.88	1.14	2.5
42	0.8	0.83	1.9
230	4.24	5.74	1.3
46	0.87	1.41	2.3
52	0.31	0.09	0.3
116	0.81	0.15	0.32
92	0.45	0.72	6.3

TABLE 7

Gene Description	5' Primer Sequence 5' to 3'	3' Primer Sequence 5' to 3'	Taqman Probe Sequence
Rat liver catalase	550	519	490
Rat Carboxylesterase	551	520	491
Rat cathepsin B	552	521	492
canalicular multidrug resistance protein	553	522	493
(s)-2-hydroxy acid oxidase	554	523	494
estrogen sulfotransferase	555	524	495
protective protein (heat shock proetin 90A	556	525	496
Rat hepatic alp-2u globulin	557	526	497
Rat transferrin	558	527	498
Cytocrhome P450	559	528	499
Aldehyde dehydrogenase, rat	560	529	500
3-methylcholanthrene-inducible UDP gluc.trans	561	530	501
rat senescence marker	562	531	502
Vitamin D binding protein, Rat	563	532	503
RB binding protein 2	564	533	
UDP-glucoronosyltransferase 1	565	534	504
mitochondrial gene fragment, Rat	566	535	505
Rat delta-aminolevulinate synthase	567	536	506
human flavoprotein	568	537	507
alpha-2u globulin, Rat	569	538	508
glutathione-S-transferase	570	539	509
rat cytosolic NADP-dependent isocitrate	571	540	510
Protein tyrosine kinase	572	541	511
hepatic steroid hydroxylase	573	542	512
Nicotinic receptor, alpha sub. unit	574	543	513
Alpha B-crystallin, heart	575	544	514
Bos Taurus aldehyde oxidase	576	545	515
lambda-crystallin	577	546	516
Vav2	578	547	517
MDM2	579	548	518
DAD1	580	549	